

Amenta, Gabriela; Vergara Rossi, Florencia. **Cómo desarrollar una biblioteca virtual con software libre: el caso de la Biblioteca Virtual para el Campus Virtual de CLACSO.** *En publicación: Babini, Dominique; Fraga, Jorge CLACSO, Consejo Latinoamericano de Ciencias Sociales, Ciudad de Buenos Aires, Argentina. 2006. pp 145-164 ISBN: 987-1183-53-4*

Disponible en la World Wide Web:

<http://bibliotecavirtual.clacso.org.ar/ar/libros/secret/babini/Amenta%20Vergara.pdf>

www.clacso.org

RED DE BIBLIOTECAS VIRTUALES DE CIENCIAS SOCIALES DE AMERICA LATINA Y EL CARIBE, DE LA RED DE CENTROS MIEMBROS DE CLACSO

<http://www.clacso.org.ar/biblioteca>

biblioteca@clacso.edu.ar

Gabriela Amenta* y Florencia Vergara Rossi** | CLACSO

Cómo desarrollar una biblioteca virtual con software libre: el caso de la Biblioteca Virtual para el Campus Virtual de CLACSO

EL CAMPUS VIRTUAL DE CLACSO¹

El Consejo Latinoamericano de Ciencias Sociales ha desarrollado durante más de una década redes destinadas a satisfacer las necesidades de los científicos sociales en América Latina y el Caribe, especialmente en el campo de la comunicación e información electrónicas. Como red de Centros Miembros, se abocó a la tarea de promover el entrenamiento de los/as investigadores/as latinoamericanos/as y caribeños/as en las nuevas tecnologías, y la creciente utilización de las redes electrónicas como instrumentos de efectiva vinculación y cooperación internacional. Es así que en 1998 se desarrolla e implementa desde el Consejo un Campus Virtual en el cual fueron alojadas las más diversas actividades académicas de CLACSO.

El Campus Virtual de CLACSO, en tanto plataforma de comunicación, información y difusión de los programas y proyectos académicos regionales e internacionales del Consejo, permite optimizar esfuerzos mediante la utiliza-

* Coordinadora del Campus Virtual de CLACSO. E-mail <gamenta@campus.clacso.edu.ar>.

** Responsable de la plataforma de la Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe, CLACSO. Diseño y desarrollo de la Biblioteca Virtual del Campus Virtual de CLACSO. E-mail <fvergara@campus.clacso.edu.ar>.

¹ Ver <<http://campus.clacso.edu.ar/>>.

ción de “espacios virtuales” para cada una de las áreas y grupos involucrados que, de este modo, han podido sostener, a lo largo del tiempo y a un bajísimo costo, un constante nivel de interacción acorde con los requisitos de una efectiva cooperación internacional. Si en el pasado esta adolecía de la inevitable intermitencia que imponían las grandes distancias existentes en la región, el funcionamiento del Campus Virtual ha hecho posible el mantenimiento de una fecunda continuidad en la labor de los académicos de la región, potenciando la interacción entre los investigadores de los centros afiliados al Consejo y facilitando enormemente la diseminación de los avances y resultados de sus actividades.

En este entorno virtual se diseñan e implementan actividades de formación a distancia y teletrabajo, conformando una gran comunidad académica donde actualmente confluyen la investigación, la docencia, el aprendizaje y la administración de espacios académicos virtuales que no podrían implementados de manera presencial. Hoy en día, integrantes de diferentes programas académicos de CLACSO hacen uso creciente de las facilidades ofrecidas por el Campus Virtual: investigadores/as ganadores/as de los diversos concursos organizados por el Programa Regional de Becas, los miembros de los Grupos de Trabajo, los/as docentes y alumnos/as del Programa de Formación a Distancia y Estudios de Posgrado. Los mismos han podido adquirir una dinámica más efectiva que en el pasado, cuando las grandes distancias y las dificultades de comunicación se erigían como obstáculos de primer orden que permanentemente menoscababan las actividades y programas académicos.

Nuestro Campus Virtual ha hecho énfasis en la calidad pedagógica, el apoyo personalizado y la utilización de instrumentos de evaluación y control de gestión, creando las condiciones para una asignación óptima de los recursos financieros y humanos, y una democratización del acceso público a los conocimientos generados por las ciencias sociales en la región. Dicha tarea mantiene sus objetivos iniciales, siendo uno de ellos la difusión y renovación del pensamiento social latinoamericano, que constituye una de las prioridades académicas de CLACSO.

Desde su formación hasta finales de 2005, en el Campus Virtual de CLACSO se han creado más de 1.800 conferencias y espacios académicos, capacitando a más de 3.700 usuarios de 32 países. Se dictaron 78 cursos de formación a distancia, una Maestría semi-presencial en Teoría y Metodología de las Ciencias Sociales y un Posdoctorado semi-presencial en Pensamiento y Cultura de América Latina, con más de 2.500 alumnos inscriptos en dichas actividades. El 57% del alumnado ha correspondido a investigadores y académicos avalados por los Centros Miembros con beca de matriculación, mientras que un 43% provino de instituciones no afiliadas al Consejo. Entre los alumnos, hubo un 47% de varones y un 53% de mujeres. El Campus Virtual también permitió teletrabajar en sus proyectos de investigación a más de 380 becarios ganadores de los concursos de becas de CLACSO, que diariamente comparten sus espacios con los tutores metodológicos a distan-

cia. Asimismo colaboró con los espacios académicos de más de mil coordinadores e investigadores de los Grupos de Trabajo del Consejo. El cuerpo académico de los cursos de formación a distancia del Campus Virtual estuvo conformado por 174 profesores, que se agruparon en 78 equipos docentes regionales e internacionales y representaron a 21 países de todo el mundo. Finalmente, a lo largo de estos años, se han diseñado, compaginado y enviado para los cursos de formación a distancia 4.138 CD-ROMs con los materiales digitalizados.

El público estudioso y especializado de las ciencias sociales ha sido el beneficiario del Campus Virtual, encontrando en el Campus un portal de acceso a resultados de investigaciones recientes, bibliografía especializada, contacto con institutos de investigación y sus investigadores, concursos de becas, y foros de debates y cursos regionales orientados a diversos campos de conocimiento en Ciencias Sociales. A su vez, las instituciones de docencia e investigación de la red de Centros Miembros de CLACSO en la región han tenido la oportunidad de difundir a través del Campus su producción científica y sus actividades, así como de brindar cursos en línea con la tecnología y metodología que ofrece el Campus Virtual. Los 32 países que se mencionan a continuación estuvieron representados a través de los alumnos en los cursos de formación a distancia: Alemania, Argentina, Australia, Bélgica, Bolivia, Brasil, Canadá, Chile, Colombia, Costa Rica, Cuba, Dinamarca, Ecuador, El Salvador, España, Estados Unidos, Guatemala, Haití, Honduras, Inglaterra, Italia, México, Nicaragua, Noruega, Panamá, Paraguay, Perú, Puerto Rico, República Dominicana, Suiza, Uruguay y Venezuela.

Desde sus orígenes, y a través del dictado de los cursos de formación a distancia en el Campus Virtual, fueron abordadas las siguientes temáticas en ciencias sociales: ALCA; Bibliotecas virtuales; Capitalismo latinoamericano; Ciudadanía y Estado; Clases sociales; Conflicto social; Cultura e identidad; Cultura y transformaciones sociales; Democracia; Derechos económicos; Derechos humanos; Derechos sociales; Desarrollo; Desarrollo urbano; Desigualdad social; Ecología política; Economía política; Economía social; Educación; Educación superior; Estructura social; Estudios del trabajo; Exclusión social; Filosofía política; Formulación de proyectos de investigación; Fuerzas Armadas; Género; Geopolítica; Globalización; Hegemonía norteamericana; Integración económica; Investigación social; Justicia social; Método Paulo Freire; Migración y modelos de desarrollo; Movimientos sociales; Movimientos sociales agrarios; Neoliberalismo; Partidos políticos; Pedagogía freireana; Pensamiento crítico latinoamericano; Pensamiento y cultura; Pobreza; Poder y nuevas experiencias democráticas; Política, mercado y sociedad; Nuevas tecnologías de información y comunicación; Políticas educativas; Políticas públicas; Políticas sociales; Praxis pedagógica de Paulo Freire; Protesta social; Pueblos indígenas; Reforma del Estado; Reforma social y económica; Relaciones internacionales; Sociedad del trabajo; Sociología de la educación;

Teoría marxista; Trabajo; Transformaciones del mundo del trabajo; Tratados de libre comercio; Violencia política; etcétera.

El cuerpo de 174 profesores que dictaron sus clases en el Campus Virtual de CLACSO y que representaron a 21 países estuvo compuesto por: Nelson Aguirre; Luis Alba; Tariq Alí; Héctor Alimonda; Clara Algranati; Sonia Alvarez Leguizamón; Elmar Altvater; Giovanni Alves; Javier Amadeo; Gabriela Amenta; Perry Anderson; Dalida Andrade; Angela Antunes; Ricardo Antunes; Enrique Arceo; Plínio de Arruda Sampaio Jr.; Dominique Babini; Carlos Barba Solano; Augusto Barrera; Juliana Barrero; Daniel Bensaid; Jaime Benvenuto Lima; Andréa Bertelli; Mikel Berraondo; Heraclio Bonilla; Paula Boniolo; Atilio A. Boron; Víctor Bretón; Irene Brouse; Pablo Cáceres; Alex Callinicos; Orlando Caputo; Guillermo Castro; Santiago Castro-Gómez; Marcelo Cavarozzi; Ana Esther Ceceña; Pablo Ciccolella; Bartolomé Clavero; Tamara Cohen Egler; Adalid Contreras Baspineiro; José Luis Coraggio; Héctor Cordone; Sonia Couto; Julio Cubillo; Pablo Dalle; Claudia Danani; Pablo Dávalos; Patricia Dávalos; David Debrott; Enrique de la Garza; Carlos Jesús Delgado Díaz; Celia Regina Días; Juan Antonio Diez; Ana Drolas; Terry Eagleton; Rodolfo Elbert; Jaime Estay; Gabriel Fajn; Ivonne Farah Henrich; Myriam Feldfeber; Mariano Feliz; Lourdes Fera Basurto; Josefina Fernández; Francisco Fernández Buey; Roberto Fernández; Roberto Fernández Retamar; Carlos Fidel; Alberto Florez; Jorge Fraga; Lutgardes Freire; Moacir Gadotti; Alicia García; Francisco García; Walter García; Pablo Gentili; Natalia Gianatelli; Esther Giraudo; Laura Golbert; Magdalena Gómez; Margarita Victoria Gómez; Alicia González; Sabrina González; Pablo González Casanova; Romeo Grompone; Luis Eduardo Grüner; María Clelia Guiñazú; Laura Gutiérrez; María Alicia Gutiérrez; Frigga Haug; Franz Hinkelammert; Susana Hintze; Silvia Korinfeld; Ivania Kuzmicic; Edgardo Lander; Verónica Lazo; Anete Brito Leal Ivo; Paula Lenguita; Roberto Leher; Bettina Levy; Alicia Lissidini; Juan Lombardo; Sinesio López; Sergio López Ruelas; Francisco López Segrera; Michael Löwy; Héctor Lucena; Alejandro Maldonado; Sara Martínez; Daniel Mato; Ellen Meiksins Wood; Elizete Menegat; Carmen Midaglia; Carmen Millán; Iliana Mignaqui; María Ángeles Monroy; Manuel Monereo; Minor Mora Salas; Manuel Moreno Castañeda; Sergio Morresi; Gustavo Navarro; Julio César Neffa; Raúl Ornelas; Pablo Ospina; Paulo Padilha; José A. Padua; Ramón Pajuelo; Héctor Palomino; Demian Panigo; Marcelo Perera; Pablo Pérez; Juan Pablo Pérez Sáinz; Ignacia Perugorrría; Octavio A. Pescador; Diego Piñeiro; Flavio Piovesan; Isabel Piper; Héctor Poggiese; Alicia Puyana; Javier Quintanilla; Estela Quintar; Rainer Randolph; Felipe Raslan; Marcos Roitman Rosenmann; Alberto Sabaté; Emir Sader; Robinson Salazar; Vania Salles; Germán Sánchez Daza; Adolfo Sánchez Vázquez; Ruth Sautu; Martín Schorr; José Seoane; Carlos Sojo; Javier Solorio; Pedro Luis Sotolongo; Ricardo Spaltenberg; Martín Spinosa; Pablo Stefanoni; Susan Street; Daniel Suárez;

Emilio Taddei; Laura Tavares Soares; Ximena Tocornal; Rafael Mario Toer; Víctor Toledo; Catalina Toro; Graciela Torrecillas; Carlos Alberto Torres; Ana Clara Torres Ribeiro; Enrique Valencia; Danilo Veiga; Rocío del Pilar Verástegui; Florencia Vergara Rossi; Ariel Wilkis; Hugo Zemelman; Raúl Zibechi; Alicia Ziccardi.

Después de varias investigaciones y pruebas de distintos tipos de software, el Campus Virtual de CLACSO optó por el FirstClass como plataforma para sus actividades académicas y estructura de comunicación e información. El mismo requirió diseños y desarrollos propios con el fin de ser personalizado y adaptado a las necesidades y características de los programas académicos del Consejo que implementarían sus actividades virtuales. El uso de esta herramienta permitió administrar con escasos recursos humanos, de manera eficiente y dinámica, los intercambios entre una gran cantidad de personas geográficamente distantes a lo largo del mundo, manteniendo los flujos de información por ellas creados. Actualmente, el Campus ofrece diversos “espacios virtuales” con mensajes de/hacia cada grupo de usuarios (docentes, alumnos, becarios, tutores, coordinadores de programas, etcétera). Estos espacios son diseñados al comienzo de cada actividad, teniendo en cuenta la necesidad detectada por el grupo electrónico como así también el vínculo de comunicación necesario para llevar adelante las actividades académicas propuestas en la región.

Por ejemplo, las aulas virtuales de los cursos de formación a distancia poseen un espacio de coordinación, otro para el dictado de las clases teóricas y la bibliografía digitalizada (obligatoria y/o de consulta), un espacio de debates y uno para la entrega de los trabajos parciales y/o finales por parte de los alumnos. Los becarios ganadores de los concursos, por ejemplo, disponen de otra complejidad en cuanto al diseño de sus espacios de interacción. En estos casos, es de vital importancia el teletrabajo con el tutor metodológico que los acompaña a lo largo de su investigación y, por otro lado, la vinculación del debate entre los becarios dedicados a las distintas temáticas con el debate intergeneracional (jóvenes, senior o semi-senior). A medida que los usuarios fueron solicitando “permanecer” en la plataforma virtual, se diseñaron nuevos espacios para la comunidad virtual, los cuales permiten mantener el vínculo de comunicación e información más allá de finalizada la actividad específica en la cual se participó (concurso de becas, curso de formación a distancia, etcétera).

LA BIBLIOTECA VIRTUAL DEL CAMPUS VIRTUAL DE CLACSO

Cada curso de formación a distancia dictado en el Campus contó con sus respectivos materiales bibliográficos que acompañaron las clases teóricas enviadas por los equipos docentes. Teniendo en cuenta el volumen de información que se encontraba digitalizado en cada CD-ROM de los cursos (más de

3.200 textos) y su calidad (manuscritos originales de prestigiosos académicos e investigadores en ciencias sociales, artículos de libros impresos, capítulos de libros agotados), se decidió, en consulta con los docentes y los responsables de la Biblioteca Virtual de CLACSO², diseñar y desarrollar la Biblioteca Virtual del Campus. Como varios de los documentos bibliográficos han sido cedidos por sus propios autores con el único objetivo de servir como material de apoyo para las clases teóricas de los cursos de formación a distancia, y dado que estos documentos tienen restricciones de *copyright* para ser publicados vía Internet, dicha biblioteca cuenta con acceso restringido para los docentes y alumnos de los cursos.

En lo que respecta al proceso de búsqueda de información en la Biblioteca Virtual del Campus, se optó por un diseño lo más amigable y familiar posible, a fin de facilitar la tarea de los alumnos y profesores que ingresan a la plataforma sin ayuda.

Históricamente, las bibliotecas se pensaron y organizaron con sus libros en los estantes (organizados por tema, colección, tipo de documento, etc.) y, en paralelo, se disponía de fichas o bases de datos ordenadas por autor, por título, por tema, etc., para el proceso de organización y búsqueda. A partir de la era digital, también podemos ofrecerle al usuario la posibilidad de consultar las versiones digitales de los textos y realizar búsquedas en el texto completo del documento. Para ello, se necesita una plataforma específica en la cual se puedan organizar los textos digitales y agregarles datos ("metadatos") que permitan realizar búsquedas en dichos metadatos (por ejemplo: autor, título, temas, editor, año, etc.), así como también en el texto completo de la publicación.

Considerando la importancia de desarrollar la Biblioteca Virtual del Campus Virtual de CLACSO en una plataforma de software gratuito de código abierto que contara con aval internacional, se optó, luego de un estudio exhaustivo de diferentes plataformas, por el software Greenstone. Greenstone es un conjunto de programas diseñado para poder crear y difundir colecciones digitales, organizar la información y publicarla en Internet o realizar CDs, con un poderoso motor de búsqueda incorporado. Es un desarrollo de la Universidad de Waikato, Nueva Zelanda, promocionado por UNESCO para el desarrollo de bibliotecas digitales. El software es de código abierto, pudiendo modificarse su apariencia y sus funciones, adaptándolas a necesidades locales. Se puede obtener en forma libre desde la página <<http://www.greenstone.org>>, y posee una Licencia Pública General de GNU (el proyecto GNU –GNU es un acrónimo recursivo para "GNU No es Unix"– comenzó en 1984 para desarrollar un sistema operativo tipo Unix completo, que fuera software libre).

2 Ver <www.clacso.org.ar/biblioteca>.

Esta plataforma “indizará” los textos que se encuentran digitalizados, en cualquiera de sus formatos (por ejemplo: PDF, WORD, RTF, etc.), convirtiéndolos al lenguaje XML³, siendo este un lenguaje muy simple, que permite al programa una gran velocidad en las búsquedas y facilita la tarea de los programas especiales cosechadores de metadatos (robots).

CARACTERÍSTICAS DE GREENSTONE

- ◆ Existe en diferentes versiones: Windows, Unix, Mac, Solaris.
- ◆ Tiene una apariencia definida que se puede modificar, por su calidad de código abierto.
- ◆ La plataforma se puede ofrecer en varios idiomas ya que es multilingüe.
- ◆ Indexa archivos de diferentes formatos, ya sean textuales (PDF, TXT, DOC, RTF) u otros (videos, fotos, bases de datos, música). Los archivos se organizan en colecciones, pudiendo crearse la cantidad de colecciones que sea necesaria.
- ◆ Cada colección se puede armar con una estructura propia.
- ◆ Trabaja con metadatos tomados automáticamente o definidos en forma manual, según el formato elegido para procesar los textos digitales.
- ◆ En el resultado de la búsqueda se visualizan los documentos del modo predefinido por el creador de la colección.
- ◆ Además de definir diferentes tipos de búsquedas, también se arman listados, con la finalidad de permitir a los usuarios navegar por la información contenida en cada colección.

3 XML es la sigla del inglés EXtensible Markup Language (lenguaje de marcado ampliable o extensible), desarrollado por el World Wide Web Consortium (W3C). Es una versión simple de SGML. Su objetivo principal es conseguir una página web más semántica. Aunque una de las principales funciones con las que nace sería suceder al HTML, separando la estructura del contenido y permitiendo el desarrollo de vocabularios modulares, compatibles con cierta unidad y simplicidad del lenguaje (objetivo que se viene desarrollando a través de la especificación XHTML), tiene otras aplicaciones entre las que se destaca su uso como estándar para el intercambio de datos entre diversas aplicaciones o software con lenguajes privados como en el caso del SOAP. Al igual que el HTML, se basa en documentos de texto plano en los que se utilizan etiquetas para delimitar los elementos de un documento. Sin embargo, XML define estas etiquetas en función del tipo de datos que está describiendo y no de la apariencia final que tendrán en pantalla o en la copia impresa, además de permitir definir nuevas etiquetas y ampliar las existentes (Wikipedia en <<http://es.wikipedia.org/wiki/XML>>).

- Desde el resultado de una búsqueda y desde los listados definidos, se accede, mediante un enlace, al texto completo.
- Indexa también bases de datos Isis y Marc.
- Tiene incorporado el sistema Open Archive Initiative⁴ (OAI) para cosecha de metadatos desde y hacia colecciones remotas en Greenstone, pudiendo así formar parte de grandes redes de información de textos digitales y bases de datos desde cada biblioteca individual.

NAVEGANDO LA BIBLIOTECA DEL CAMPUS VIRTUAL DE CLACSO

Desde la Biblioteca del Campus Virtual se puede tener acceso a los textos mediante las funciones de búsqueda y consulta.

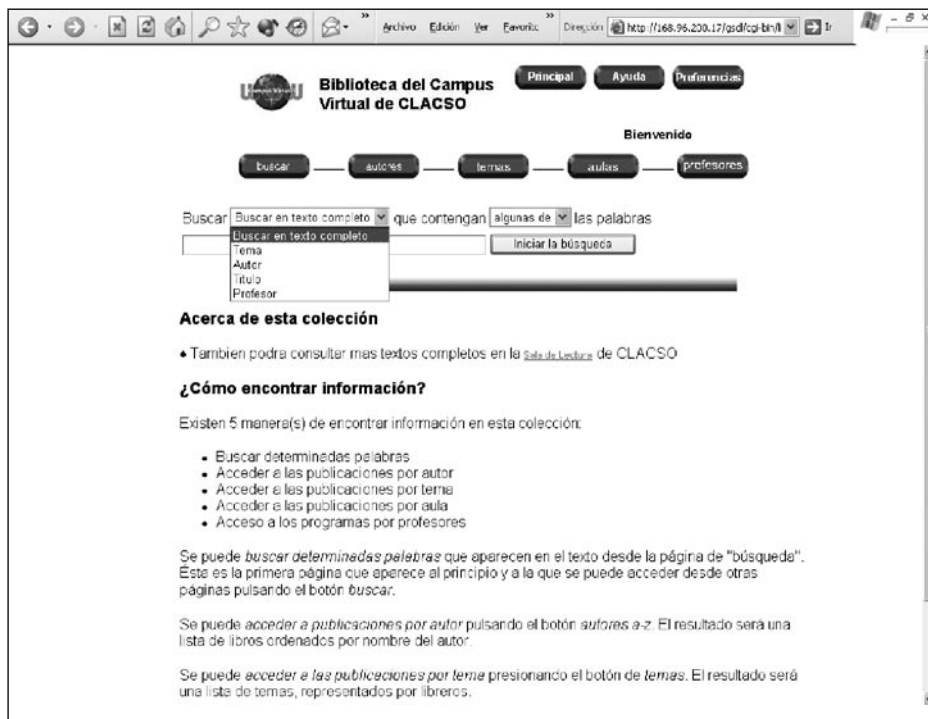
BÚSQUEDA

En la búsqueda, el programa Greenstone examina el texto completo de todos los documentos incluidos en la biblioteca y sus respectivos metadatos.

Las posibilidades que se definieron en la función de búsqueda para la Biblioteca Virtual del Campus Virtual de CLACSO son las siguientes:

- búsqueda en texto completo
- por tema
- por autor
- por título
- por profesor

⁴ Ver <www.openarchives.org>.



Es muy importante elegir previamente los metadatos a incluir en cada texto, ya que la plataforma tomará la información de estos para armar las diferentes funciones de búsquedas y de consulta.

En la Biblioteca Virtual del Campus Virtual de CLACSO se utilizan los siguientes metadatos de los 15 básicos Dublin Core⁵, además de metadatos propios.

- ◆ Title: Título del capítulo
- ◆ Autor: Autor del capítulo
- ◆ Description_2: Título del libro o publicación seriada donde se encuentra el capítulo
- ◆ Creator: Compilador
- ◆ Publisher: Editorial

⁵ Ver <<http://dublincore.org>>.

- 📌 Date: Año
- 📌 Month: Mes
- 📌 Pages: Páginas del artículo y/o capítulo
- 📌 Aula: Aula en donde se consignó el texto como bibliografía
- 📌 Copy: *Copyright* del texto
- 📌 Profesor: Profesor del Campus Virtual que asignó este texto en su bibliografía del curso
- 📌 Insti: Institución donde realiza sus actividades el profesor
- 📌 Subject: descriptores temáticos para clasificar el texto
- 📌 Subject_1: descriptor general para ese texto, con la finalidad de poder agruparlo en estantes temáticos

CONSULTA

Otra posibilidad para consultar los textos incluidos en la biblioteca es hacerlo por medio de listados especialmente definidos, con índices alfabéticos ordenados de diferentes formas. Los listados definidos en esta biblioteca son:

- 1 Listado de Autores: es un listado de los autores de los textos. Cuando hay más de un texto del mismo autor, se agrupan en un estante virtual, que al desplegarse mediante un clic mostrará los textos existentes de ese autor. El listado está ordenado alfabéticamente.
- 2 Listado de Temas: dentro de los metadatos definidos para esta colección, existe el metadato de temas (Subject, Subject_1), tanto para poder realizar búsquedas por descriptores de los textos, como para poder además agruparlos en estantes virtuales clasificados por temas.
- 3 Listado de Aulas: se agrupan los textos que se han utilizado en cada curso impartido en el Campus Virtual de CLACSO. El listado de textos dentro de cada aula está ordenado alfabéticamente por autor del texto.
- 4 Listado de Profesores: listado alfabético de los profesores que han dictado sus clases en los distintos cursos. Se accede al programa de es-

tudios del curso dictado por cada profesor, con el listado de los textos exigidos para el curso.



CÓMO ARMAR UNA COLECCIÓN: LA EXPERIENCIA DE LA BIBLIOTECA DEL CAMPUS VIRTUAL DE CLACSO

Como ya se mencionó, Greenstone procesa diferentes formatos de archivos de contenido textual, como los textos digitalizados para los diferentes cursos del Campus, principalmente en formatos PDF y DOC. Antes de ingresar los textos a Greenstone, que los transforma en formato XML, se los procesa manualmente en formato HTML⁶. La utilización de este formato permite:

- ◆ colocar metadatos manualmente;
- ◆ con los metadatos, definir automáticamente las citas bibliográficas para cada texto, como un servicio adicional para el usuario;

6 Con los textos en formato PDF o DOC, sólo se podrán armar listados por título dentro de la colección, pues es el único metadato que Greenstone podrá identificar como exacto dentro del texto, ya que el proceso es automático.

- ✿ definir índices temáticos de la bibliografía: el usuario podrá navegar dentro de la biblioteca por diferentes listados ordenados alfabéticamente;
- ✿ especificar tipos de búsqueda por distintos metadatos: autor, título, tema, profesor;
- ✿ colocar una tapa a cada texto.

PROCESANDO TEXTOS EN HTML, FORMATO ELEGIDO PARA TODOS LOS TEXTOS

Para cada texto a procesar, se necesitan tres archivos: HTML, METADATA.XML y JPG.

- ✿ En el archivo HTML estará el texto, con su estructura y sus respectivos metadatos.
- ✿ En el archivo METADATA.XML estarán sólo los metadatos del texto, que se asociarán con el HTML.
- ✿ El archivo JPG será la tapa del texto (libro, revista, etcétera).

En la Guía del Programador, en el sitio web de Greenstone⁷, se explica cómo armar un texto en el formato HTML, trabajando con su código en el bloc de notas o en el *word pad*, armando estructuras determinadas en secciones; se indican además las razones para crear secciones y subsecciones en los archivos HTML (pp. 20-21).

El lenguaje HTML necesita etiquetas que definan cada paso del diseño del texto, tales como un título “en negrita”, los espacios entre párrafos, etcétera. Las etiquetas HTML correspondientes a las diferentes funciones del lenguaje generalmente se colocan en forma manual, y además hay que completar los correspondientes metadatos. El costo en tiempo es bastante alto, motivo por el cual en CLACSO hemos desarrollado una herramienta especial llamada Pi05.

VENTAJAS DEL PI05

- ✿ Facilitar el trabajo, agilizar el procesamiento de los textos
- ✿ Colocar etiquetas HTML estándar automáticamente
- ✿ Definir las secciones en el texto, con sus metadatos

⁷ Ver <www.greenstone.org> versión 2.60.

- ✦ Generar un documento HTML y el archivo METADATA.XML
- ✦ Funcionamiento en forma local y en línea
- ✦ Colocar enlaces, insertar imágenes y realizar el diseño del texto fácilmente
- ✦ Software bajo licencia GNU-GPL
- ✦ Posibilidad de asociar la tapa del texto: archivo JPG

El programa cuenta con los grupos de metadatos ya predefinidos para los diferentes textos a procesar: un libro, una publicación periódica, un artículo, una tesis.

PASOS A SEGUIR

- ✦ abrir el texto original: PDF, DOC, RTF, etcétera;
- ✦ copiar el texto y pegarlo en un bloc de notas o word pad y guardarlo en formato TXT;
- ✦ luego abrir el programa Pi05 y, desde el programa, abrir el texto TXT que acabamos de generar;
- ✦ completar los metadatos del texto, incluir la tapa (archivo JPG) y guardar. Para obtener una explicación ampliada y acceder al programa libremente, ingresar en la página <<http://www.clacso.org.ar/biblioteca/pi05>>.

CONFIGURACIÓN DE GREENSTONE PARA ARMAR LA BIBLIOTECA VIRTUAL

A modo de ejemplo para quienes estén interesados en desarrollar bibliotecas virtuales, nos referiremos específicamente a la configuración de la colección para la Biblioteca Virtual del Campus Virtual de CLACSO. Todas las variaciones se podrán ver en la Guía del Programador en el sitio web de Greenstone, y en colecciones ya creadas por otras instituciones, que se podrán utilizar a modo de ejemplo.

Primero es necesario instalar el programa. A tal fin, se recomienda ver la guía de Instalación de la Biblioteca Digital Greenstone. Allí se explican las instrucciones para todo el proceso para cada sistema, y también cómo comenzar una sesión en Greenstone. Para comenzar a trabajar la plataforma, el programa debe estar instalado y funcionando con las colecciones de ejemplo incluidas en Greenstone al instalarlo.

A fin de crear nuestras propias colecciones una vez instalado el programa Greenstone, la versión de Greenstone para Windows incluye dos interfaces para poder crear y armar colecciones paso a paso: el Collector y la Interfaz

del Bibliotecario. En CLACSO realizamos el proceso en forma manual, desde la consola de MS-DOS. Los comandos se explican en la Guía del Programador de Greenstone (pp. 2-3).

Cuando los comandos se trabajan manualmente, los textos que se procesan se van acumulando, y se actualiza la colección en forma más rápida; no se guardan los originales dentro de la colección. En cambio, si utilizamos la Interfaz del Bibliotecario, cada vez que agreguemos nuevos textos toda la información será procesada nuevamente, y deberemos tener siempre los originales dentro de la colección. Los originales se colocan en la carpeta *import* de la colección.

En CLACSO utilizamos la versión de Greenstone para Linux, pues el servidor funciona en este sistema. Y para realizar pruebas, utilizamos la versión en Windows de manera local. La Interfaz del Bibliotecario sólo funciona en Windows.

CREANDO UNA COLECCIÓN

Una vez que los textos se van procesando en HTML, con sus metadatos correspondientes ingresados, armamos nuestra colección desde la línea de comandos, con las instrucciones para armar colecciones que se encuentran en la guía del Programador de Greenstone (pp. 2-3).

Cuando una colección ha sido creada, se genera la siguiente estructura de carpetas:

◆ ARCHIVES

◆ BUILDING

◆ ETC

◆ IMAGES

◆ IMPORT

◆ INDEX

◆ PERLLIB

En la carpeta ETC encontramos un archivo llamado CONFIG.CGF. Lo abrimos en un bloc de notas o en *word pad*. Es aquí donde cambiaremos y agregaremos las líneas de configuración de la colección. Al abrir el archivo, aparece una configuración estándar generada automáticamente, que vamos a personalizar.

Cada atributo aporta una parte de información sobre la colección que determina su apariencia y el modo en que se van a tratar los documentos.

A continuación se mostrará el modo en que se definió la Biblioteca del Campus Virtual de CLACSO. Cada línea de atributos se encuentra numerada para facilitar la explicación.

ARCHIVO CONFIG.CGF DE LA BIBLIOTECA DEL CAMPUS VIRTUAL

- 1 creator vergara@clacso.edu.ar
- 2 maintainer vergara@clacso.edu.ar
- 3 public true
- 4 indexes section:text document:Subject section:Autor section:Title
section:Profesor
- 5 plugin ZIPPlug
- 6 plugin GAPlug
- 7 plugin TEXTPlug
- 8 plugin HTMLPlug -description_tags -cover_image -nolinks
- 9 plugin EMAILPlug
- 10 plugin PDFPlug
- 11 plugin RTFPlug
- 12 plugin WordPlug
- 13 plugin PsPlug
- 14 plugin ArcPlug
- 15 plugin RecPlug -use_metadata_files
- 16 classify AZCompactList -metadata Autor -sort dls.Autor -buttonname
Creator
- 17 classify AZCompactList -metadata Subject_1 -sort Creator -mingroup
1 -buttonname Subject
- 18 classify AZCompactList -metadata dls.Organization -sort Autor -
buttonname Organization
- 19 classify AZCompactList -metadata Profesor -sort Profesor -buttonname
Countries
- 20 format SearchVList " <td valign=top>[link][icon]/[link]</td><td>{lf}{[par
ent(All\':\'):Title], [parent(All\':\'):Title]:} [Autor] [link][Title]/[link]{lf}{[dls.
Publisher],
Editor: [Publisher]} {lf}{[dls.Mounth],[Mounth]}{lf}{[
dls.Date],Fecha: [Date]}{lf}{[dls.Description_1],
En publicacion:
[Description_1]}{lf}{[dls.Pages],[Pages]} {lf}{[dls.Aula],
[Aula]}{lf}{[dls

- .Copy],
[Copy]]{If}{[dls.Profesor],
[Profesor]]{If}{[dls.Insti],
[Insti]]{If}{[dls.Course],<i>[Course]</i>}{If}{[dls.Program],
[link][Program]/link]]<p></td>"
- 21 format CL1VList "<td valign=top>[link][icon]/link</td><td valign=top>{If}{[numleafdocs],[Title],[link][Autor]/link
[Title]
En libro: [Description_1]
Editor: [Publisher]. Fecha: [Date]<small>
[Aula]</small><p></td>"
- 22 format CL2VList "<td valign=top>[link][icon]/link</td><td valign=top>[highlight]{Or}{[Autor]}/highlight<small>{If}{[Title],
[link][Title]/link} {If}{[Publisher],
Editor: [Publisher]} {If}{[Mounth],
[Mounth]}{If}{[Date], [Date]}{If}{[Description_1],
En libro: [Description_1]}{If}{[Aula],
Aula: [Aula]} {If}{[Subject_1],
Temas: [Subject_1]}{If}{[Copy],
[Copy]}</small><p></td>"
- 23 format CL3VList "<td valign=top>[link][icon]/link</td><td valign=top>[highlight]{Or}{[Autor]}/highlight<small>{If}{[Title],
[link][Title]/link} {If}{[Publisher],
Editor: [Publisher]} {If}{[Mounth],
[Mounth]}{If}{[Date],
[Date]}{If}{[Description_1],
En libro: [Description_1]}{If}{[Aula],
Aula: [Aula]}{If}{[Copy],
[Copy]}</small><p></td>"
- 24 format CL4VList "<td valign=top>[link][icon]/link</td><td valign=top>{If}{[numleafdocs],[Title],[link][Profesor]/link- <small><i>[Course]</i>. [Insti] </small><p></td>"
- 25 format DocumentText "[Autor]. [Title]. {If}{[Description_1],En libro: [Description_1].}{If}{[Publisher],[Publisher].}{If}{[Mounth],[Mounth]}{If}{[Date], [Date]}{If}{[Aula],
Aula: [Aula]}<p>{If}{[Subject], Descriptores: [Subject]}-{If}{[Subject_1], [Subject_1]}\n\n<p>[Text]"format DocumentImages true
- 26 format HList"[link][highlight]{Or}{[dls.Title],[dc.Title],[Title],Untitled}/highlight]/link]{If}{[Description_1],
En Libro: [Description_1]}"
- 27 collectionmeta collectionname "Biblioteca del Campus Virtual de CLACSO"
- 28 collectionmeta iconcollection "_httpprefix_/collect/bcvirt/images/campus.jpg"
- 29 collectionmeta .section:text [l=es] "Buscar en texto completo"
- 30 collectionmeta .document:Subject [l=es] "Tema"
- 31 collectionmeta .section:Autor [l=es] "Autor"

32 collectionmeta .section:Title [l=es] "Título" collectionmeta .section:
Profesor [l=es] "Profesor"

33 # -- Spanish text ----- collectionmeta
collectionextra [l=es] " ♦ Tambien podra consultar mas
textos completos en la Sala de
Lectura de CLACSO<p>"

SIGNIFICADO DE CADA LÍNEA

- 1 Creador: dirección electrónica del creador de la colección
- 2 Maintainer: correo electrónico del actualizador de la colección
- 3 Public: determina si la colección va a ponerse a disposición del público o no
- 4 Indexes: definimos los tipos de búsquedas y los niveles en esa colección:
Section:text document:Subject section:Autor section:Title section:Profesor.

Los comandos indican que se podrá buscar en el texto completo, por tema (descriptores, subject), autor, título, profesor. Si tuviéramos un libro completo con capítulos, cada capítulo se llama y define como sección (section) y lo que indicamos es que las búsquedas se realicen en los metadatos de cada sección.

Para ampliar los distintos niveles de índices consultar la Guía del Programador Greenstone (indexes), pp. 23.

5-15. CONECTORES (PLUGINS)

Los conectores analizan los documentos y extraen los metadatos de los textos digitales. El conector HTML extrae los metadatos que están entre etiquetas, que este lenguaje utiliza. Al procesar el texto en Pi05, los datos ingresados en cada campo (título, autor, etc.) quedan insertos automáticamente en dos etiquetas HTML. Estas etiquetas contienen el nombre del metadato y la información, por ejemplo: <Metadata name="Title">El título de algo</Metadata>, y así para cada uno de los metadatos que completamos. La información que se encuentra en cada metadato la "tomará" el conector (plugin) HTML para armar las búsquedas que definamos, los listados y el resultado de dichas búsquedas, según lo precisemos.

Los conectores que el programa necesita siempre son: GaPlug, PsPlug, ArcPlug, RecPlug (ver especificaciones técnicas la Guía del Programador, pp. 29). A los conectores se les pueden agregar opciones.

- ◆ **description_tags:** interpreta los archivos de documentos etiquetados. Es específico de este conector. Toma las secciones asignadas en los documentos.

- ❖ **cover_images:** busca un archivo JPG (con el mismo nombre que el archivo HTML) y lo asocia con el documento siendo la tapa del texto.
- ❖ **nolinks:** no guarda los enlaces dentro de la colección; reconoce los enlaces como externos.
- ❖ **use_metadata_files:** asocia el archivo METADATA.KML con el archivo HTML

16-19. CLASIFICADORES

Los clasificadores sirven para armar diferentes listados que acompañan a la búsqueda. Por ejemplo, armar un listado con los títulos de la colección ordenados alfabéticamente. Sirven para crear los índices de consulta de la colección: autores, temas, aulas, profesores (Guía del Programador de Greenstone, pp. 37-44).

- ❖ Todos tienen una línea que empieza por la palabra clave *classify*, seguida del nombre del clasificador y las opciones que admite.
- ❖ Classify AZList –metadata Title crea una lista alfabética de títulos reuniendo todos aquellos documentos que están provistos del campo de metadatos Title, y los clasifica y divide por orden alfabético.

ARGUMENTOS DEL CLASIFICADOR

- ❖ Metadata: identifica los metadatos mediante los cuales se clasifican y seleccionan los documentos. El clasificador omitirá todo documento de la colección para el que no se haya definido este metadato (pero se indiza, y el buscador lo tendrá en cuenta para las búsquedas). Si no se especifica un metadato, todos los documentos se incluyen en el clasificador.
- ❖ Buttonname: define el botón de consulta de Greenstone que activa el clasificador.
- ❖ Cada clasificador recibe un nombre implícito con arreglo a su posición en el archivo de configuración. Por ejemplo: el tercer clasificador especificado en el archivo se denomina CL3. Estos nombres se utilizan para nombrar los campos de la base de datos de informaciones de la colección que definen la jerarquía de los clasificadores.

VEAMOS CÓMO CONFIGURAR A CADA UNO

LISTADO DE AUTORES

```
16 classify AZCompactList -metadata Autor -sort dls.Autor -buttonname  
Creador
```

LISTADO TEMÁTICO

17 `classify AZCompactList -metadata Subject_1 -sort Creador -mingroup 1-buttonname Subject`

LISTADO DE AULAS

18 `classify AZCompactList -metadata dls.Organization -sort Autor -buttonname`

ORGANIZATION

LISTADO DE PROFESORES

19 `classify AZCompactList -metadata Profesor -sort Profesor -buttonname Countries`

Cada parte de la sentencia significa

- 🔹 `Classify`: siempre.
- 🔹 `AZCompactList`: crea un listado alfabético. Cuando la información del metadato `Autor` coincide, la agrupa generando un estante dentro del listado.
- 🔹 `metadata Autor`: se define qué metadato tomará el sistema para armar el listado. En este caso será el metadato `Autor`, pues es el listado que nos interesa armar.
- 🔹 `mingroup1`: arma estantes automáticos del metadato asignado, en este caso de temas, agrupando los textos.
- 🔹 `sort dls.Autor`: Aquí le indicamos al sistema por cual de los metadatos se debe ordenar el listado a generar. También elegimos el de `Autor`.
- 🔹 `buttonname Creator`: indica qué botón utilizará el programa para acceder al listado. Son botones existentes en Greenstone.

20 Definimos cómo se presentarán los datos en el resultado de una búsqueda.

`Format SearchVList`: define de qué manera se presentarán los datos en el resultado de una búsqueda

- 🔹 `format`: cadena de formatos en donde se especifica cómo se presentarán los datos en el resultado de una búsqueda, y los listados que deseamos crear para que se pueda navegar en la colección y ver los contenidos desde ahí.

- 21-24 Establecemos cómo se presentarán los datos en cada listado que definimos. `Format CL1VList`: define cómo se verán los datos de los textos en los diferentes listados; el `CL1VList` corresponde al listado de títulos, el `CL2VList` corresponde al listado de autores.
- 25 Definimos el encabezado en cada texto. En el caso de la Biblioteca del Campus Virtual se definió la cita bibliográfica, que se arma automáticamente, tomando la información de los metadatos de cada texto.
- 27 `Collectionname`: es el título que queremos que aparezca en la colección.
- 28 `Collectionicon`: si en vez de una frase queremos insertar un logo en cada colección, desde aquí indicaremos la ruta y el nombre del archivo del gráfico, escribiendo la siguiente sentencia con el nombre del gráfico; el gráfico lo colocaremos en la carpeta `IMAGES` dentro de la colección.
- 29-32 `Collectionmeta`: define los tipos de búsqueda que ofrecemos al usuario; indicamos el metadato: qué metadato utilizará la plataforma, y la frase que debe aparecer en pantalla.
- 33 `Collectionextra`: Greenstone incluye la posibilidad de colocar información en cada colección, que se denomina "Acerca de esta colección". Se podrá escribir un párrafo, insertar imágenes, crear tablas, colocar enlaces, describir la colección, etcétera.

Cuando hemos terminado de armar el archivo de configuración de la colección, `CONFIG.CFG`, lo guardamos en la carpeta `ETC`. Agregaremos los textos ya procesados en `HTML`, para que la colección comience a funcionar.

INDEXANDO LOS TEXTOS

En la carpeta `IMPORT` copiaremos la carpeta que hemos generado con el `Pi05` con los tres archivos: `HTML`, `METADATA.XML` y `JPG`.

Desde la línea de comandos procesaremos los textos de la siguiente manera: primero realizar un `import.pl`, que codifica los textos en lenguaje `XML`, y luego realizar un `buildcol.pl`, que arma los listados definidos en cada colección, según se detalla en la Guía del Programador (pp. 8-10).

Una vez que se han realizado los procesos `import.pl` y `buildcol.pl`, en la carpeta `BUILDING` quedarán creadas varias carpetas. En Windows se deberán copiar estas carpetas generadas en la carpeta `INDEX`, porque desde allí el programa buscará la información.

Finalizado este proceso, al abrir la Biblioteca Virtual Greenstone, se verá el resultado de nuestra creación.